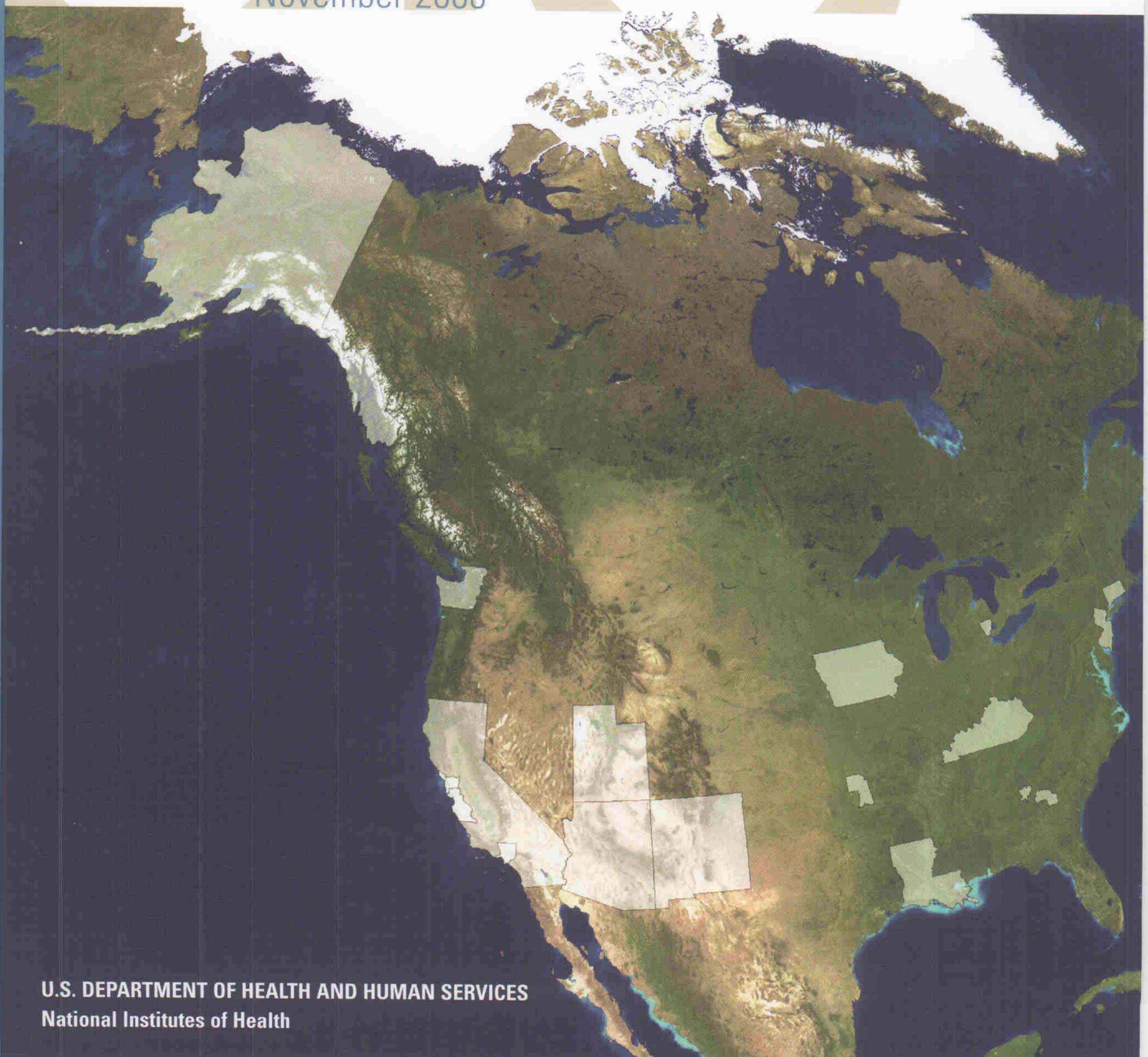
National Cancer Institute

Division of Cancer Control and Population Sciences

Geographic Information Systems (GIS) and Cancer Research

November 2006



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

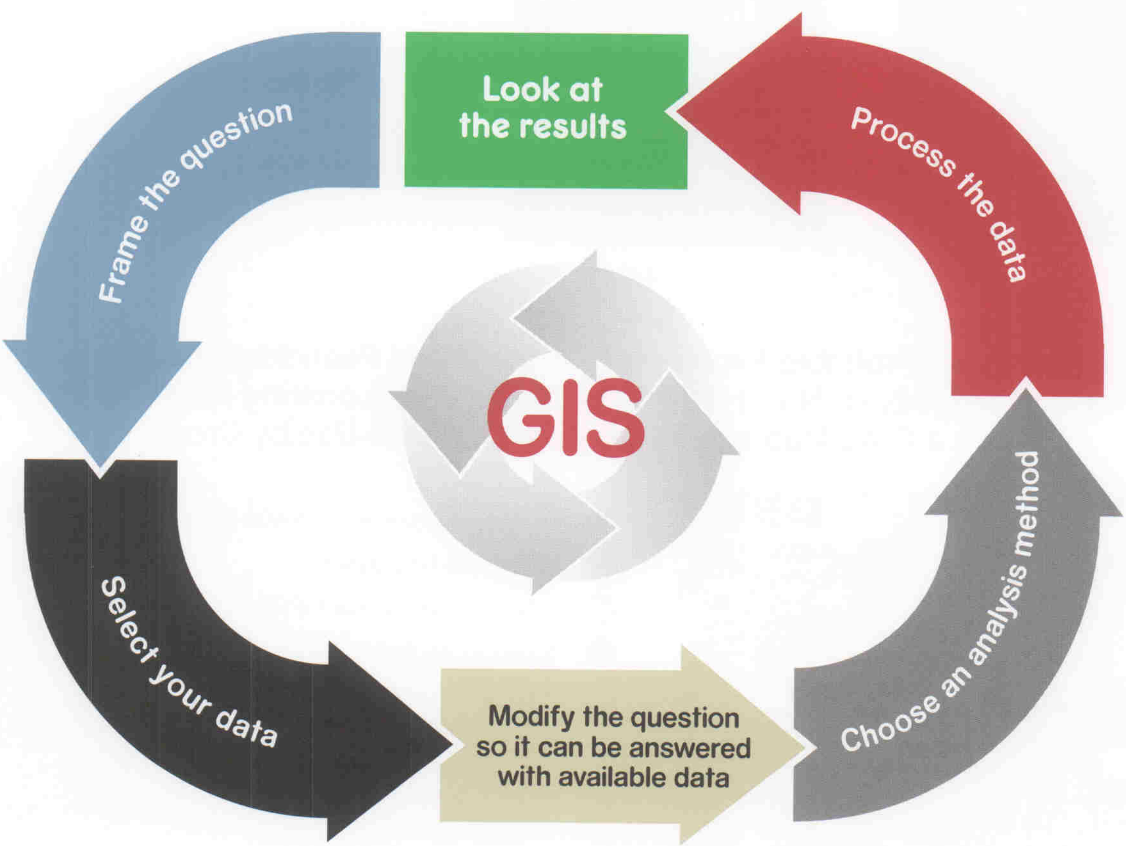
Investigating geographic patterns of cancer in the United States for more than 30 years—from epidemiologic research to identify causes of cancer in local areas, to studying the impact of personal behavior and community characteristics on health care outcomes, to targeting cancer prevention activities where they are needed most.

Use GIS in Your Work

Geospatial tools are used for a variety of applications to

- Identify and display the geographic patterns of cancer incidence and mortality rates in the United States and their change over time
- Create complex databases for the study of cancer screening, diagnosis, and survival at the community level
- Assess environmental exposure
- Estimate cancer incidence, prevalence, and survival for every U.S. state, using spatial statistical models
- Communicate local cancer information to the public and public health professionals through interactive Web-based tools
- Expose health disparities at the local level through comparison of cancer outcomes across demographic subgroups
- Develop new methods of displaying geospatial data for clear communication to the public and for examination of complex multivariate data by researchers

Answering Questions with GIS



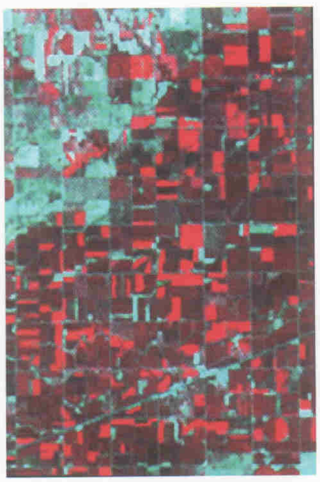
Adapted from Lance Waller, Emory University, and John Richardson, EPA

Applications for GIS Technology in Specific Areas of Research

- **Environmental Exposure Assessment**
  A GIS can provide information about potential environmental exposures that cannot be obtained through traditional epidemiologic methods such as questionnaires.

  For example, a study in south central Nebraska demonstrated use of satellite imagery to reconstruct historical crop patterns, which were then used to estimate an individual's potential exposure to specific agricultural pesticides.

Constructing Historical Crop Patterns, South Central Nebraska*



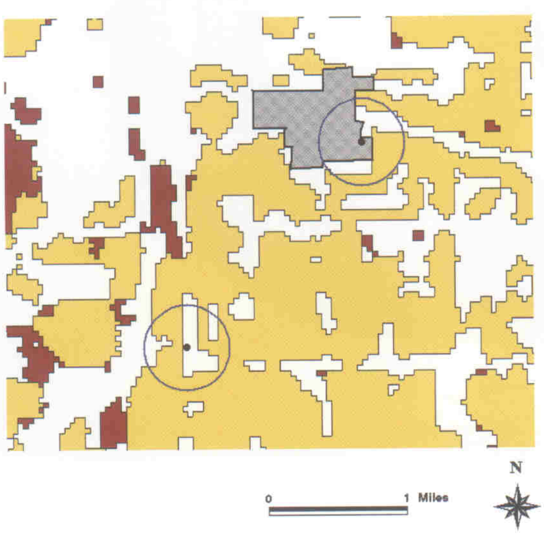
Landsat Color Infrared Imagery

Classified Land Cover Map

Classified Land Cover Map
Class Name

- range/pasture/grass/cut alfalfa
- fallow/bare soil/road
- corn
- sorghum
- soybeans
- alfalfa-full cover
- urban
- water

Estimating Probable Exposure to Agricultural Pesticides in a Case-Control Study of Non-Hodgkin Lymphoma by Locating Residences on a Crop Map and Estimating Pesticide Use by Crop*



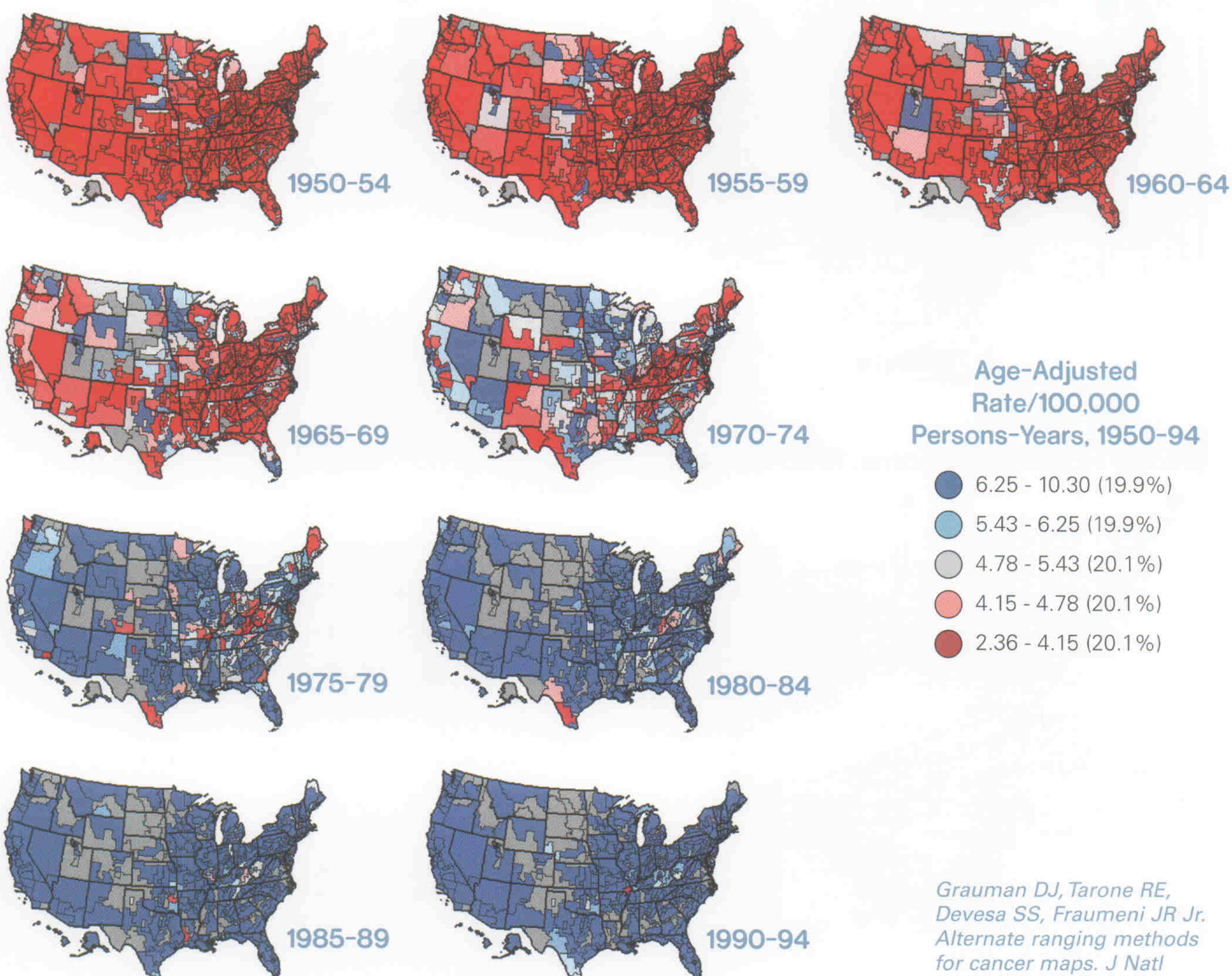
- Individual residences with 500-meter buffer
- USGS place boundaries
- Atrazine applied to corn (Prob. = 0.54)
- Propachlor applied to sorghum (Prob. = 0.54)
- Other land cover types

**Ward MH, Nuckols JR, Weigel SJ, Maxwell SK, Cantor KP, Miller RS. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System. Environ Health Perspect. 2000 Jan;108[1]:5-12*

In an ecologic study of prostate cancer mortality rates (1950-2000) in four northern plains states (Minnesota, Montana, North Dakota, and South Dakota), investigators found that rates were higher in rural counties compared with urban counties and that higher rates were significantly associated with crop production. Specifically, prostate cancer mortality rates increased as the percentage of population living within 500 meters of small grains crops increased.

- **Cancer Control**
  GIS technology allows researchers to monitor emerging trends in our national cancer burden and the factors that influence these measures, such as screening availability and utilization. The information gathered through GIS technology helps to guide the cancer control research agenda and allows for appropriate decision making by identifying opportunities for investment with high payoff in terms of reduced morbidity and mortality. For example, the figure below shows that although U.S. cervical cancer mortality has dropped dramatically, high rates persisted longer in some areas, such as Appalachia and the Texas-Mexico border area.

Changing Patterns Over Time: Cervical Cancer Mortality among White Females, 1950-94

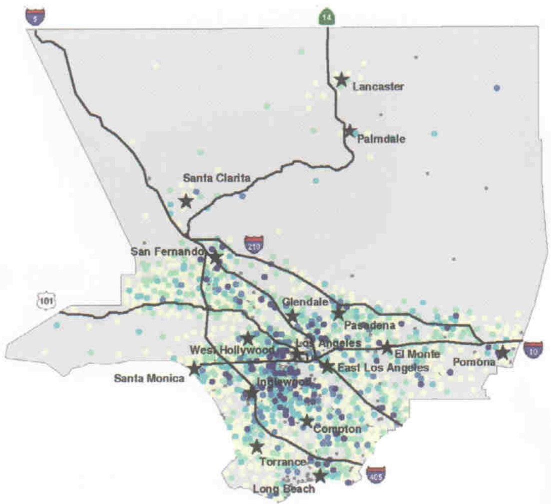


Grauman DJ, Tarone RE, Devesa SS, Fraumeni JR Jr. Alternate ranging methods for cancer maps. J Natl Cancer Inst 2000;92:534-43.

- **Health Disparities**
  A study of breast cancer cases diagnosed at a late stage in Los Angeles county showed striking geographic differences among the county's census tracts. These patterns were strongly associated with tumor histology as well as income, marital status, and distance to the nearest mammography facility from within their tract of residence.

Percent Advanced Disease among Breast Cancer Cases Diagnosed in Los Angeles Census Tracts, 1992-96, Compared with Median Income

Percent Advanced Disease

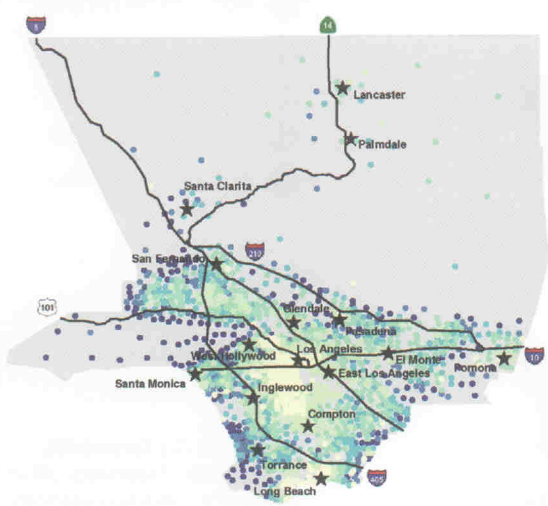


Percent

- 0% - 5%
- 5% - 10%
- 10% - 15%
- 15% - 20%
- > 20%

Darker color = higher % advanced disease

Median Household Income, 1990 Census



Income

- Under $20,000
- $20,000 - $30,000
- $30,000 - $40,000
- $40,000 - $50,000
- $50,000 - $60,000
- Over $60,000

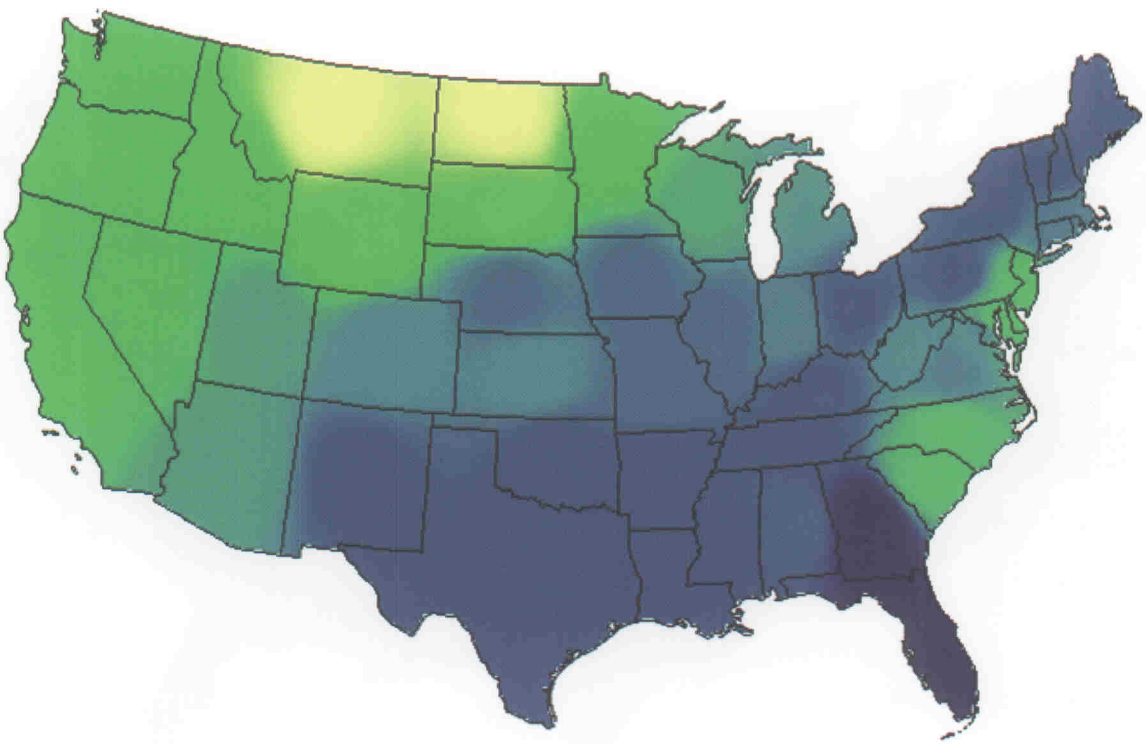
Darker color = higher median income

Gumpertz ML, Pickle LW, Miller BA, Bell BS. Geographic patterns of advanced breast cancer in Los Angeles: associations with biological and sociodemographic factors (United States). Cancer Causes Control. 2006 Apr;17(3):325-39.

• **Behavior Research**

The Health Information National Trends Survey (hints.cancer.gov) tracks trends in the public's use of new communication technologies while charting progress in meeting health communication goals in terms of the public's knowledge, attitudes, and behaviors. Using GIS technology, such as weighted smoothing and surface interpolation, researchers are able to display and compare the distribution of cancer risk factor knowledge according to geographic area, providing valuable information for further health communication research and health promotion programs, policies, and practices.

Knowledge Map—Sun Exposure and Cancer Risk



Does exposure to the sun increase a person's chances of getting cancer a lot, a little, or not at all?

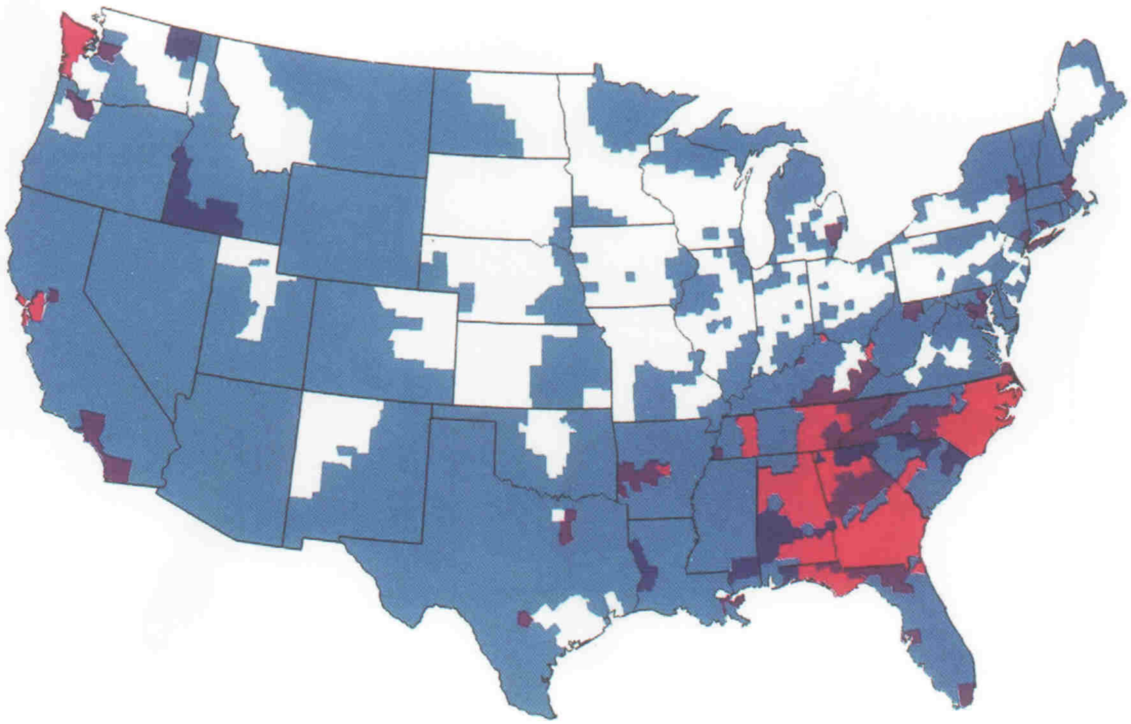
Percent Responding "A-Lot"

Data from 2003 Health Information National Trends Survey (hints.cancer.gov)

http://gis.cancer.gov

- **Health Policy**
  A landmark NCI study published in 1981 found that an oral cancer cluster was due to snuff dipping (smokeless tobacco). The findings led to a law requiring warning labels on smokeless tobacco packaging and printed advertisements, and a ban on electronic advertising. (Winn DB et al. N Engl J Med. 1981; also see page 13)

Oral Cancer Mortality, 1950-69, by State Economic Areas



Age-Adjusted Rate

- Significantly higher than U.S. In highest decile
- Significantly higher than U.S. Not in highest decile
- In highest decile. Not significant
- Not significantly different from U.S.
- Significantly lower than U.S.

Mason TJ, McKay FW, Hoover R, Blot WJ, Fraumeni JF Jr. Atlas of cancer mortality for U.S. counties: 1950-1969. Washington, DC: U.S. Gov. Printing Office; 1975. DHEW Publ. No. (NIH) 75-780.

GIS Tools and Resources

GIS Databases

- **GIS for Breast Cancer Studies on Long Island**
  The Geographic Information System for Breast Cancer Studies on Long Island (LI GIS) is a GIS comprising data with statistical and spatial extensions. The LI GIS is designed to study the potential relationships between environmental exposures and breast cancer in Nassau and Suffolk counties, New York. The LI GIS also can be used to study other diseases. More than 80 datasets covering topographic, demographic, health outcome, and environmental data are included.

  Researchers may apply online to use the LI GIS and/or the LI GIS statistical software and a full suite of GIS software with spatial extensions. Access is free. An interactive mapping system plus GIS software extensions related to the study of breast cancer are available.
  www.healthgis-li.org

  The LI GIS is one of a series of initiatives within the Long Island Breast Cancer Study Project (LIBCSP), a congressionally mandated activity to understand breast cancer incidence on Long Island.

> **Applications:** ***Access data, interactive mapping systems, and customized software to investigate relationships between cancer and the environment***

- **TOXMAP**
  TOXMAP, developed by the National Library of Medicine, is a user-friendly interactive tool that uses maps to explore the Environmental Protection Agency's (EPA) Toxics Release Inventory. Data found in TOXMAP come from several providers, including

  —Toxics Release Inventory (TRI)
  —EPA Comprehensive Environmental Response, Compensation and Liability Information System (CERCLIS)
  —EPA Facility Registry System (FRS)
  —National Library of Medicine Hazardous Substances Databank® (HSDB)
  —National Library of Medicine TOXLINE®
  —Centers for Disease Control and Prevention (CDC) Agency for Toxic Substances and Disease Registry (ATSDR)
  —National Weather Service OST/SEC GIS Map Group
  —ESRI and Geographic Data Technology, Inc.
  —GeoGratis

  TOXMAP is a free service, which can dynamically generate specific custom-made TRI maps via simple links; there are no restrictions on and no limit to the number of custom maps that can be created.
  toxmap.nlm.nih.gov

> **Applications:** ***Generate custom maps and visually explore environmental data representing multiple, integrated data sources***

- **SEER Database**
  The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) is an authoritative source of information on cancer incidence and survival in the United States. SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 26 percent of the U.S. population. The SEER Program provides data free of charge through reports on cancer statistics and in a public-use data file. Software is also provided free of charge by NCI that aids in the analysis of the SEER database (SEER*Stat) and can be used by other cancer registries to analyze their data (SEER*Prep and SEER*Stat).
  seer.cancer.gov

> **Applications:** ***Build statistical profiles using cancer incidence and survival data covering approximately 26 percent of the U.S. population to identify and evaluate geographic and population differences in cancer patterns and other socio-demographic factors***

Statistical Methods

- **Statistical Modeling**

  Statistical models provide a means to identify significant associations between potential cancer risk factors and cancer incidence, mortality, prevalence, and other cancer outcomes. Spatial models take into account the similarity of cancer rates in neighboring places. Results of these models can provide better estimates of cancer rates in places with small populations, or can estimate the cancer rate or burden in areas without a cancer registry.

  For example, hierarchical Poisson regression models were used to characterize associations between cancer incidence and mortality, sociodemographic, and lifestyle factors by county. The models provided an estimated number of new cancer cases by state and county for combinations of cancer site, gender, age, and year.

  Model results have been used to

  —Predict the number of new cases expected in the coming year by state

  —Predict cancer prevalence

  —Calculate the percent completeness of case ascertainment for each U.S. cancer registry

- **Cluster Identification**

  Are apparent map clusters real or chance occurrences? To help answer that question, SaTScan software is used to identify the most likely significant cluster over space, time, or both. Within moving scan windows of varying size, the observed number of cases is compared with the expected number of cases, and the likelihood of this clustering occurring by chance is evaluated. The software has recently been extended to clusters of survival rates. **www.satscan.org**

  SaTScan is designed for any of the following interrelated purposes:

  —Evaluate reported spatial or space-time disease clusters, to see if they are statistically significant

  —Test whether a disease is randomly distributed over space, over time, or over space and time

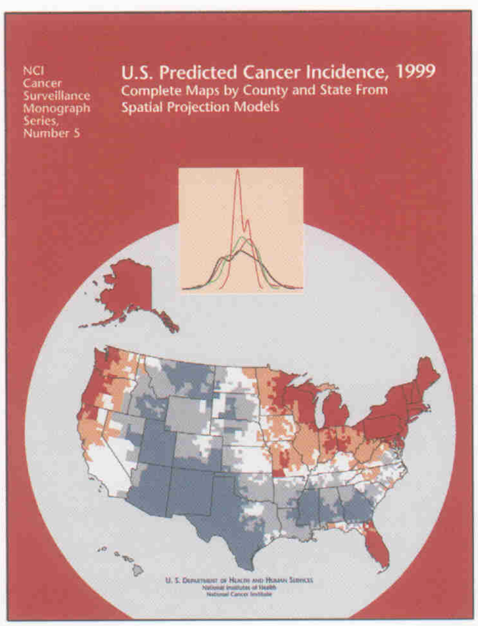
  —Perform geographical surveillance of disease, to detect areas of significantly high or low rates

  —Conduct periodic disease surveillance for the early detection of disease outbreaks

  —SaTScan can be applied to Poisson counts with an underlying population at risk or to Bernoulli or binomial event data, such as case-control data

(Developed by Martin Kulldorff. Stat Med, 1995; Commun Stat, 1997; Am J Epidemiol, 1997; Am J Public Health, 1998. [Also see pages 13 and 14]).

U.S. Predicted Cancer Incidence, 1999: Complete Maps by County and State from Spatial Projection Models



http://srab.cancer.gov/incidence

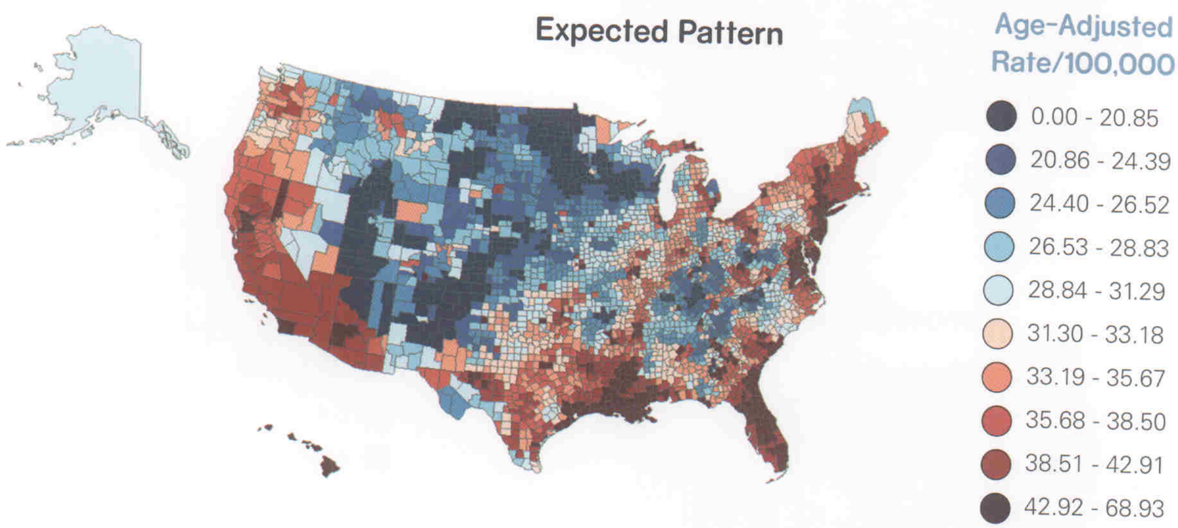
Applications: ***Explore geographic patterns of cancer incidence and mortality and study the impact of cancer control interventions by using statistical models to estimate cancer rates, both historic and projected into the future***

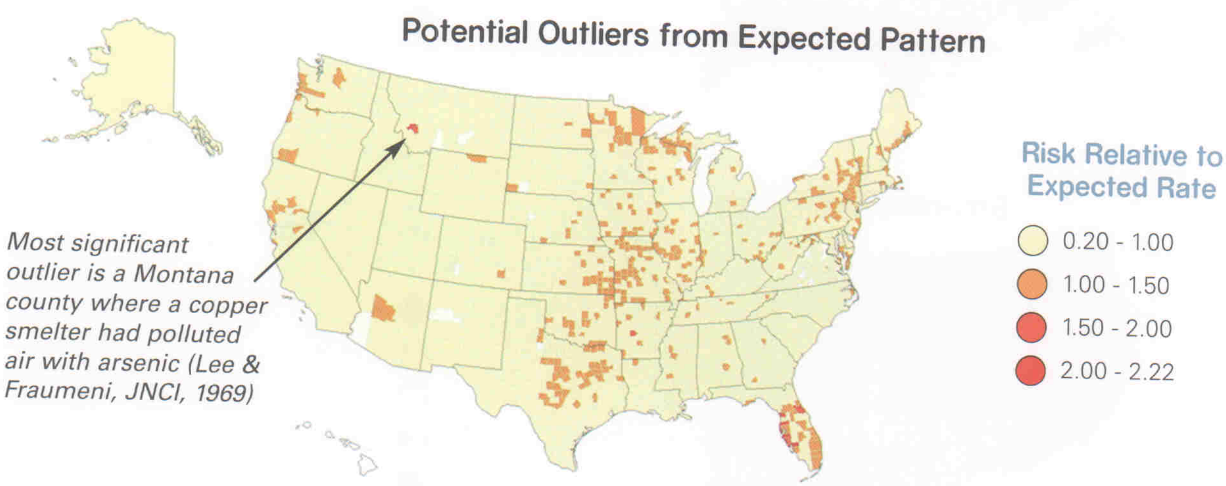
- **Outlier Detection for Cancer Surveillance**
  Once we understand the usual spatial patterns for a type of cancer, can we detect significant outliers (unusual occurrences) to these patterns when new cancer case counts are available? An outlier detection tool, such as the multi-item gamma Poisson shrinker (an empirical Bayes data mining tool [DuMouchel and Pregibon, 2001]), can be applied to both historical and more recent cancer mortality data sets, which can be stratified by cancer site, gender, and race/ethnicity.

  The tool identifies individual geographic locations with higher than expected event counts, or "hot spots," and can be used to also detect possible local emerging hot spots.

Applications: ***Analyze recent and historical data through alternative methods to detect possible local emerging hot spots***

Detection of Significant Outliers in Male Lung Cancer Mortality, 1950-69, Using an Empirical Bayes Tool





Pickle L, et al. Geographic-Based Research and Applications at the National Cancer Institute [poster]. GIScience 2002, Boulder, Colorado.

http://gis.cancer.gov

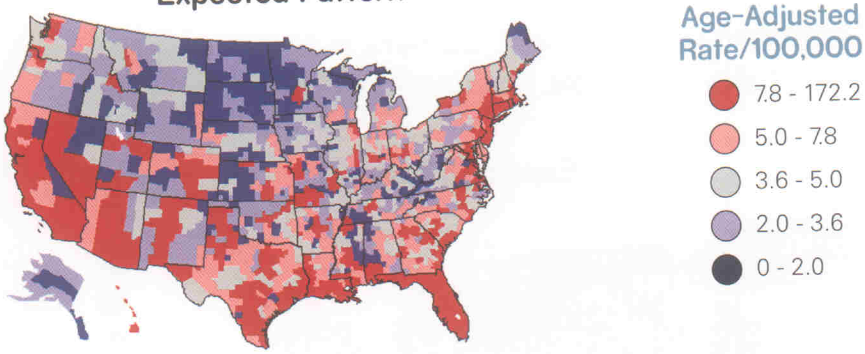
Geovisualization Tools

NCI staff have collaborated with university researchers to develop new graphical methods to display cancer statistics and their spatial patterns, for example

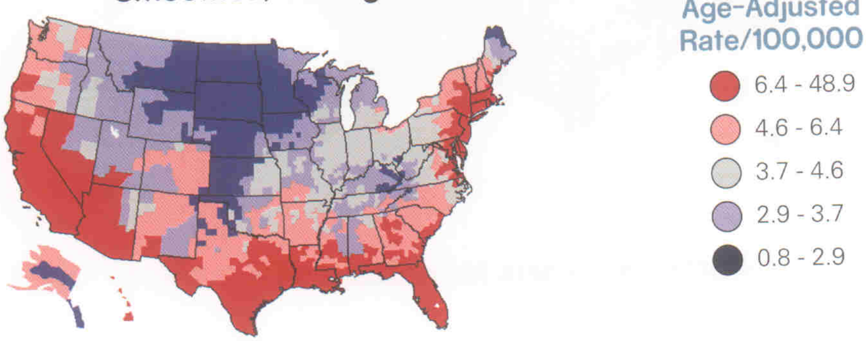
- Weighted Nonparametric Smoothing ("Headbanging")—smoothes less reliable rates more than very reliable ones
- Linked Micromap Plot—ties statistical graphics to maps by the use of color

HIV Mortality Rates, 1988-92

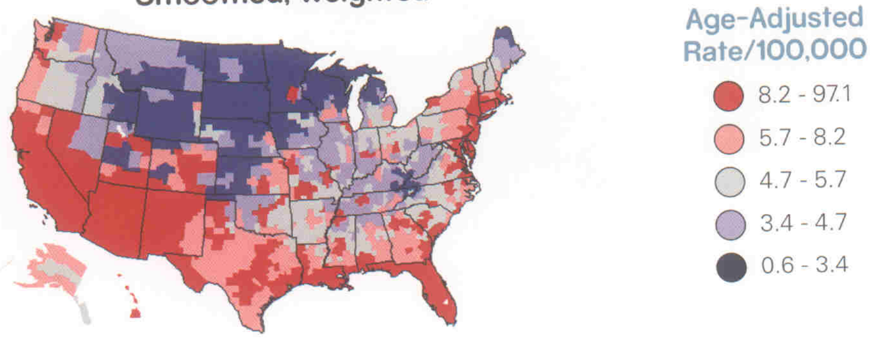
Expected Pattern



Smoothed, unweighted



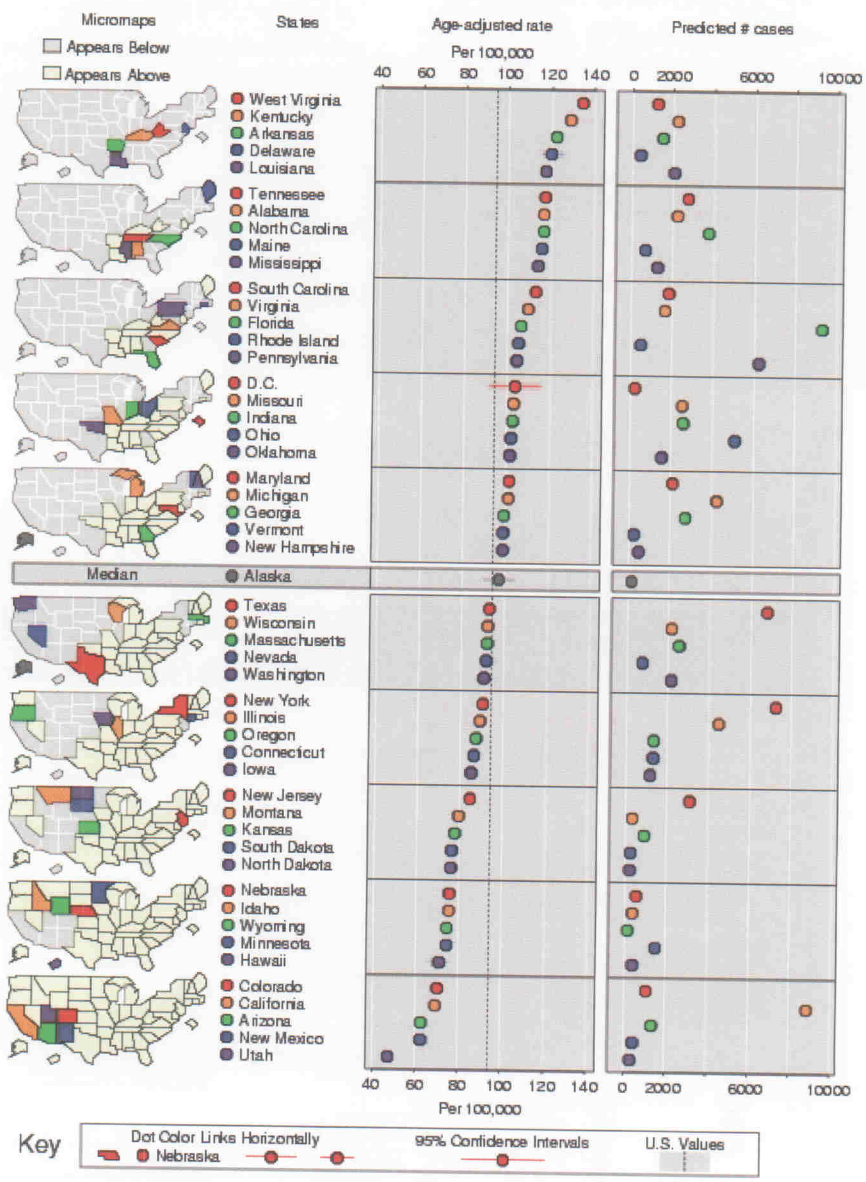
Smoothed, weighted



This is an active area of research with many more tools under development.

Applications: ***Explore the spatial and statistical structure in data through a visual application to promote innovative analysis***

Linked Micromap Plot: Linking Age-Adjusted Rates and Predicted Case Counts for Male Lung Cancer, 1999



http://srab.cancer.gov/incidence

http://gis.cancer.gov

Communication of Georeferenced Statistics

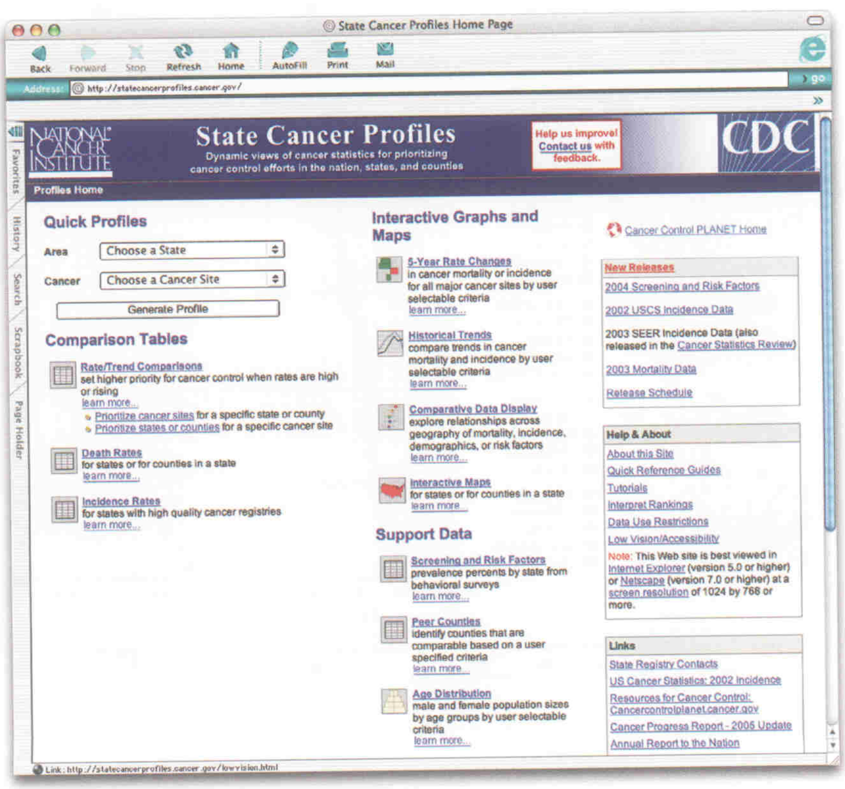
- **State Cancer Profiles**
  State Cancer Profiles is a comprehensive system of interactive maps, graphs, and tables enabling the investigation of cancer trends at the national, state, and county level. Data sources include NCI's Surveillance, Epidemiology, and End Results (SEER) Program, Centers for Disease Control and Prevention (CDC), and National Center for Health Statistics (NCHS). The objective of the State Cancer Profiles Web site is to provide a system to characterize the cancer burden in a standardized manner in order to motivate action, integrate surveillance into cancer control planning, characterize areas and demographic groups, and expose health disparities. The focus is on cancer sites for which there are evidence-based control interventions, and the interactive graphics and maps provide visual support for deciding where to focus cancer control efforts.
  statecancerprofiles.cancer.gov

Applications: ***Retrieve state and county-specific comparative incidence and mortality data to further cancer control planning***

- ***Atlas of Cancer Mortality in the U.S., 1950-94***
  The Atlas of Cancer Mortality in the U.S., 1950-94, available online, provides dynamic, interactive maps, text, tables, and figures showing geographical patterns and time trends of cancer five-year death rates from 1950 to 1994 for more than 40 cancers. In addition, the data used in creating the maps, charts, and graphs on the Web site may be downloaded.
  www3.cancer.gov/atlasplus

Applications: ***Find insights into the potential causes of cancer by exploring the geographic patterns of mortality rates***

http://statecancerprofiles.cancer.gov



Landmark GIS Studies

1969—Arsenic from Copper Smelters Leads to Lung Cancer

An area of an unusually high rate of lung cancer mortality in Montana led to the identification of arsenic exposures among copper smelter workers as the cause. (Lee AM, Fraumeni JF Jr. Arsenic and respiratory cancer in man: an occupational study. J Natl Cancer Inst. 1969 Jun;42[6]:1045-52.)

1975—Snuff Dipping Leads to Cancer

The first NCI cancer atlas identified a striking cluster of high rates of oral cancer among white women living in the southeast. Winn and colleagues pinpointed snuff dipping (smokeless tobacco) as the cause. This landmark study led to a later ban on the sale of smokeless tobacco to minors. (Winn DB, Blot WJ, Shy CM, Pickle LW, Toledo A, Fraumeni JF Jr. Snuff dipping and oral cancer among women in the southern United States. N Engl J Med 1981;304[13]:745-749.)

1978—Asbestos in World War II Shipyards Leads to Lung Cancer

Exceptionally high rates of lung cancer were seen in several areas of the east coast of the United States during the 1950s. Investigations of these hotspots identified the cause as asbestos exposure from World War II shipyards. (Blot WJ, Harrington JM, Toledo A, Hoover R, Heath CW Jr, Fraumeni JF Jr. Lung cancer after employment in shipyards during World War II. N Engl J Med. 1978;299[12]:620-4.)

1997—Breast Cancer: Persistent Northeast Excess

As a demonstration of the newly released cluster detection program SaTScan, Kulldorff et al. analyzed breast cancer mortality patterns in the United States. Although it was well known that rates had been high in northeastern states for decades, this program found an area stretching from New York City to Philadelphia that had the highest rate. (See page 14.) (Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: A geographic analysis. Am J Epidemiol 1997;146:161-170.)

2005—Built Environment Leads to More Walking

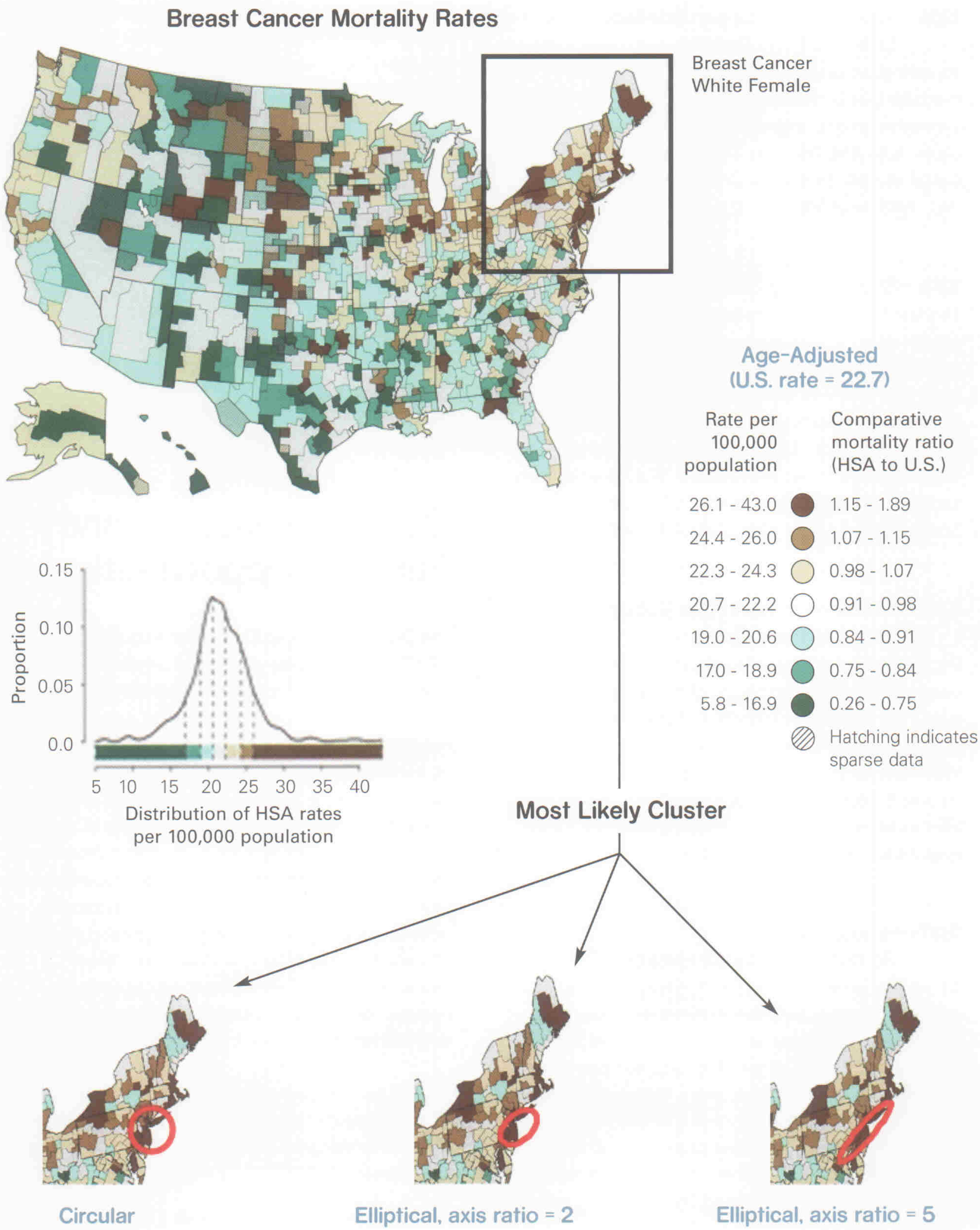
Because of a growing interest in associations between energy balance and carcinogenesis, NCI researchers are analyzing associations between street connectivity (for example, contrasts between inner city and suburban cul-de-sac rich communities), neighborhood characteristics such as income and land use characteristics, and walking for multiple purposes. Additionally, the investigators are using techniques developed to find cancer clusters for identifying areas with increased walking prevalence. (Berrigan D, Dill J, Adamski R, Huang L, Stinchcomb D, Davis W, Pickle L. National Cancer Institute, Bethesda, MD; Portland State University, Portland OR. The built environment and walking/bicycling for transportation: two approaches to identifying environmental correlates of behavior. Sixth International Conference on Dietary Assessment Methods, April 27-29, 2006, Copenhagen 2006.)

Current Research and Funding Opportunities

The Division of Cancer Control and Population Sciences (DCCPS) funds a large portfolio of grants and contracts. The portfolio currently includes over 900 grants valued at almost $400 million. The breadth of research supported by DCCPS includes surveillance, epidemiology, health services, behavioral science, and cancer survivorship. The division also plays a central role within the federal government as a source of expertise and evidence on issues such as quality of cancer care, the economic burden of cancer, geographic information systems (GIS), statistical methods, communication science, tobacco control, and the translation of research into practice. The intramural Division of Cancer Epidemiology and Genetics uses GIS to conduct etiologic investigations designed to evaluate the role of environmental factors in the development of cancer.

DCCPS currently supports over 20 grants and contracts on GIS research, valued at approximately $4 million (see **gis.cancer.gov**). It also hosts a sabbatical program for professors working on spatial statistical models and geovisualization.

Example: Breast Cancer Clusters



Below is information on current funding opportunities that may be relevant to GIS research. DCCPS participates in many other funding opportunities that are listed on the division Web site. Information about all of the funding opportunities below, how to apply for them, and on NIH resources for applicants can be found on the DCCPS Web site.
cancercontrol.cancer.gov

Small Business Grants

http://www.cancercontrol.cancer.gov/funding_sbir.html

Small businesses may obtain support through the Small Business Innovation Research (SBIR) and the Small Business Technology Transfer Research (STTR) Programs. These programs are designed to support innovative research that has the potential for commercialization.

Exploratory Grants for Behavioral Research in Cancer Control (R21) PA-06-351

This exploratory grant is a two-year award designed to encourage exploration of new ideas and methodologies in the target area and to provide support for the collection of pilot data to be used as the basis for later R01s. Studies may focus on

1. Assessment
2. Intervention
3. Dissemination
4. Surveillance
5. Biological and psychological influences on cancer and cancer-related behaviors

Cancer Surveillance Using Health Claims-Based Data System (R21, R01) PA-06-385 and PA-06-386

Cancer surveillance may include assessment of patterns of care, quality and outcomes of care, and health disparities across the continuum of treatment. Projects sought under this announcement may focus on treatment and outcomes at the patient-specific level or include influences from the provider or broader health-system level. Applications may also initiate analyses to expand understanding of the capability of and methods needed to use claims data for cancer surveillance.

Who We Are (Leadership)

The National Cancer Institute (NCI) has had an active research program investigating geographic patterns of cancer in the United States for more than 30 years. These activities are conducted in collaboration with researchers across NCI, at other federal agencies, and in academia.

- DCCPS's Statistical Research and Applications Branch (SRAB) contributes to GIS research by developing statistical methods for the analysis, display, and Web-based communication of geo-referenced cancer data. Examples of GIS work done and publications written by SRAB can be found at **srab.cancer.gov/about/areas/gis_spatial.html**.

- DCCPS's Cancer Statistics Branch (CSB) oversees the Surveillance, Epidemiology, and End Results (SEER) Program (**seer.cancer.gov**), an integrated, comprehensive, multiple population-based reporting system of cancer registries covering 26 percent of the U.S. population. Cancer incidence information is provided at the state, county, and census tract level. Investigators in SEER's Rapid Response Surveillance Studies (RRSS) are exploring and applying GIS technology in several areas.
seer.cancer.gov/rapidresponse

- The Epidemiology and Genetics Research Program (EGRP) of DCCPS has funded the development and maintenance of the Geographic Information System for Breast Cancer Studies on Long Island (LI GIS). The LI GIS is available to researchers and can be used to study other types of cancer and conditions as well. EGRP also supports research that applies GIS techniques to improve understanding of cancer risks.

- DCEG's Occupational and Environmental Epidemiology Branch has an active research program in using GIS, satellite imagery, environmental databases, and other resources to estimate potential exposure to cancer-causing environmental agents based on residential histories of participants in environmental epidemiology studies conducted by DCEG. Exposures evaluated include agricultural pesticides, arsenic and nitrate levels in private wells, and dioxins and furans from incinerators and industrial sources.

- DCEG's Epidemiology and Biostatistics Program has a long history of publishing GIS data at the county level through the NCI cancer atlases. The NCI atlas published in 1975 was the first look at the geographic patterns of cancer mortality at the county level. The maps generated many hypotheses about the causes of cancer and led to more than 100 epidemiologic studies across the U.S. Results from these studies included such important findings as the link between occupational exposure to asbestos and lung cancer in port cities of World War II, snuff dipping and oral cancer in southeastern states, and increased lung cancer mortality in Montana due to environmental arsenic pollution. The more recent atlas published in 1999 and subsequent analyses revealed substantial geographic, temporal, gender, and racial variations in lung cancer mortality related to cigarette smoking patterns, in cervix uteri cancer mortality related to socioeconomic and health care patterns, breast cancer mortality and risk factor profiles, and latitude and sunlight exposure associated with melanoma and other skin cancer but not non-Hodgkin lymphoma. The observation of elevated bladder cancer mortality rates among white men and women persisting over several decades stimulated the NCI New England Bladder Cancer Study to evaluate several hypotheses, including the potential role of arsenic in drinking water from wells.

NCI staff serve as advisors for geographic projects at the U.S. Geologic Survey, National Science Foundation, National Academy of Sciences, New York State Health Department, Long Island Breast Cancer Study Project, California Health Interview Survey, and the GIS Committee for the North American Association of Central Cancer Registries.

Future of GIS

- Satellite imagery will be used increasingly for health research. Ward, Nuckols, and colleagues demonstrated the utility of this resource by estimating probable pesticide exposure from satellite images of crop fields, something not estimable by any other method.

- Neighborhood characteristics have been found to impact health behaviors, access to care, and other health-related factors. Hierarchical analyses, including neighborhood as well as individual effects, will become increasingly important for cancer studies. Additional statistical methods for hierarchical models will permit the analysis of these complex data.

- The continued improvements in computer technology will make possible more powerful geovisualization tools for exploring and displaying complex layers of geographic information.

- Better tools and methods will be available for identifying, handling, and communicating spatial uncertainty.

- An emphasis on community-based participatory research in health studies will provide a richer set of local geographic information for analysis.

- More data will be readily available through Web-based portals, facilitating analyses of data from a wider variety of sources.

Additional Information and Resources

For more information about GIS at NCI, please visit http://gis.cancer.gov


http://gis.cancer.gov

Geographic Information Systems (GIS) and Cancer Research

Front cover: A map of Surveillance, Epidemiology, and End Results (SEER) surveillance areas draped over a cloudless composite satellite image from NASA's "Visible Earth" program.

National Cancer Institute

NIH Publication No. 07-6096
Printed November 2006

T 0 8 7